Video Article

Biochemical and Structural Characterization of the Carbohydrate Transport Substrate-binding-protein SP0092

Simone Culurgioni[1,2], Minzhe Tang[1,2], David R. Hall[1], Martin A. Walsh[1,2]

[1]Diamond Light Source, Harwell Science & Innovation Campus

[2]Research Complex at Harwell, Harwell Science & Innovation Campus

Correspondence to: Simone Culurgioni at simone.culurgioni@diamond.ac.uk

URL: https://www.jove.com/video/56294
DOI: doi:10.3791/56294

Keywords: Biochemistry, Issue 128, ABC transporters, *Streptococcus pneumoniae*, carbohydrate uptake, substrate-binding protein, thermal shift assay, multi angle light scattering (MALS), size exclusion chromatography (SEC), crystallography, X-ray diffraction, data collection, protein structure solution

Date Published: 10/2/2017

Citation: Culurgioni, S., Tang, M., Hall, D.R., Walsh, M.A. Biochemical and Structural Characterization of the Carbohydrate Transport Substrate-binding-protein SP0092. *J. Vis. Exp.* (128), e56294, doi:10.3791/56294 (2017).

Abstract

Development of new antimicrobials and vaccines for *Streptococcus pneumoniae* (pneumococcus) are necessary to halt the rapid rise in multiple resistant strains. Carbohydrate substrate binding proteins (SBPs) represent viable targets for the development of protein-based vaccines and new antimicrobials because of their extracellular localization and the centrality of carbohydrate import for pneumococcal metabolism, respectively. Described here is a rationalized integrated protocol to carry out a comprehensive characterization of SP0092, which can be extended to other carbohydrate SBPs from the pneumococcus and other bacteria. This procedure can aid the structure-based design of inhibitors for this class of proteins. Presented in the first part of this manuscript are protocols for biochemical analysis by thermal shift assay, multi angle light scattering (MALS), and size exclusion chromatography (SEC), which optimize the stability and homogeneity of the sample directed to crystallization trials and so enhance the probability of success. The second part of this procedure describes the characterization of the SBP crystals using a tunable wavelength anomalous diffraction synchrotron beamline, and data collection protocols for measuring data that can be used to resolve the crystallized protein structure.

Video Link

The video component of this article can be found at https://www.jove.com/video/56294/

Introduction

S. pneumoniae (pneumococcus) is a gram-positive bacterium residing asymptomatically in the upper airways of the human respiratory tract with the ability to migrate to normally sterile niches causing otitis, pneumonia, sepsis, septicemia, and meningitis[1,2]. Moreover, pneumococcal infection is the leading cause of community-acquired pneumonia, which is contributing to a clinical and economic burden worldwide[3,4]. Antibiotic resistant strains of *S. pneumoniae* have spread across the globe and although a seven-valent and a thirteen-valent pneumococcal protein conjugate vaccine have helped reduce the rate of antimicrobial resistance, replacement strains from vaccine use have emerged and have led to increased demands for research into the development of new treatments for pneumococcal disease[5,6,7,8].

The pneumococcus depends on sugars imported from the host as a carbon source[9,10]; indeed it devotes 30% of its import machinery to the transport of 32 different carbohydrates[11,12,13]. These importers include at least eight ABC-transporters[13]. In ABC transporters, the extracellular SBPs play a fundamental role in determining the specificity for the ligand and presenting it to the integral membrane transporter for uptake in the cell. SBPs represent valid targets for the design of new vaccines and antimicrobials because they are surface proteins and their vital role in cellular processes.

Target protein characterization and detailed description of structural features, like ligand pockets and interdomain flexibility, provide a useful tool for structure-based drug design[14,15]. X-ray crystallography is the method of choice for the structural characterization of proteins at near to atomic resolution, but the crystallization process is unpredictable, time-consuming, and not always successful. Systematic methods have improved the success rate and important factors are the sample quality and stability. The success rate of crystallization is influenced by the protein chemical properties and sample preparation methodology. The effect of these can be assessed and informed by biochemical characterization[16,17].

A further complication for structure-based design is the crystallographic phase problem, which must be addressed. As more protein structures have become available, many structures can be resolved by the molecular replacement method, which requires a homologous structure[18]. As SBPs present a flexible domain structure, molecular replacement may also prove challenging[19]. If a structural model that is sufficiently similar to the target protein is not available, a number of techniques can be used to obtain the experimental phasing[20]. Among these, the Single-wavelength Anomalous Dispersion (SAD) method has emerged as the primary technique and has been extensively used to solve the phase problem[21]. The use of the SAD method has been further advanced with improvements in hardware and software, as well as data collection strategies to allow the detection and use of weak anomalous signals for phasing[22,23,24]. Furthermore, advances in direct methods for solving

structures of macromolecules, which in the past required diffraction data to atomic resolution, can now be utilized by combining for example, stereochemical knowledge as implemented in the program ARCIMOLDO[25]. A useful review of methods for solving the phase problem in crystallography is given by Taylor[26].

Here we present a rationalized protocol for the characterization of the carbohydrate transport SBP, SP0092 of *S. pneumoniae*, integrating biochemical and structural techniques (**Figure 1**). This step-by-step protocol provides a useful example test case of strategies to improve the success rate of structural studies on SBPs in general, which are found in all kingdoms of life. In particular, the protocol highlights the importance of characterizing the most stable oligomeric state of the protein in solution in a fast and effective method, and allows the identification of the best species to follow up for crystallization experiments. Although there are over 500 SBP structures reported in the Protein Data Bank[27], molecular replacement can be challenging due to the inherent flexible nature of the two α/β domains, which are connected by a hinge region[19]. Thus, the second part of the protocol describes using the SAD method for phasing from bound metal ions, which is common in SBPs, as well as the incorporation of selenomethionine and use of selenium (Se) in SAD phasing.

Protocol

NOTE: The coding sequence for which the signal peptide is deleted is cloned in the pOPINF vector following a standard in-fusion protocol; the native protein is expressed as a His-tag fusion in *Escherichia coli* BL21 Rosetta cells[28,29]. The selenomethionine labeled variant is expressed following standard methods according to the manufacturer[30]. The recombinant SBP is purified as previously described[31,32].

1. Biochemical Characterization

1. Thermal shift assay
    1. Prepare 48 buffer solutions with a pH ranging from 4.0 to 9.5 and NaCl concentration from 0 to 0.5 M, and dispense 40 µL in a 96-well plate as described in **Table 1**[33].
    2. Add to each well 5 µL of the SBP solution at 1 - 2 mg/mL concentration.
    3. Add to each well 5 µL of 20x fluorescent stain for proteins (see **Table of Materials**).
    4. Mix the solutions by pipetting, seal the plate using a transparent adhesive film, and spin for 2 min at 112 x g at room temperature.
    5. Run the experiment on a real-time machine: Set a temperature ramp of 3 °C/min from 4 to 99 °C with a 10 s hold time, and read the fluorescence every 0.5 °C with excitation at 483 nm and emission at 568 nm.
    6. Analyze the fluorescence emission curves for every buffer condition, identifying the melting temperature (T_m) as the minimum of the derivative.
    NOTE: This procedure gives a good indication of the best buffer solution to aid in improvement of protein stability. Choose the pH range and salt concentration based on the buffer condition with the highest T_m and add 2.5% (v/v) of glycerol and 0.5 mM of tris(2-carboxyethyl)phosphine (TCEP) to define the buffer solution for the following steps (SEC-buffer).

2. MALS and analytical SEC
    1. Perform a two-column volume wash of the pre-packed gel filtration column with degassed filtered water (use a wide range molecular weight 24 mL column).
    2. Connect the column to the light scattering detector and equilibrate the column with the determined SEC-buffer from the thermal shift assay.
    3. Inject 100 µL of purified SBP at 5 mg/mL into the equilibrated SEC column and flow one column volume of SEC-buffer.
    4. Check the elution chromatogram for single or multiple peaks and examine the scattering data using the analysis software, obtaining the molar mass and the polydispersity index for all species.
    NOTE: For every peak corresponding to an oligomeric species, it is beneficial to check the polydispersity index calculated as molar mass weighted on mass/molar mass weighted on number of molecules (Mw/Mn). A polydispersity value close to 1 indicates a monodispersed peak.
    NOTE: Once all the oligomerization species are defined, it is useful to study their dependence on protein concentration, as higher concentrations might favor larger oligomer formation.
    5. Execute multiple analytical SEC runs; inject 100 µL of purified SBP at increasing concentrations (0.1 - 10 mg/mL) into the equilibrated gel filtration column (use a wide range molecular weight 5 mL column) and flow one column volume of SEC-buffer each time.
    6. Check the elution chromatogram for single or multiple peaks and analyze the absorbance intensities at 280 nm of the curves corresponding to different starting protein concentrations. If the relative intensities of the different peaks remain constant, then no inter-conversion between the oligomeric species is present. Variation in the relative intensities at different concentrations is an indication of the interconversion between different oligomeric species that is dependent on the protein concentration.
    NOTE: This characterization step is fundamental to define which species are more suitable for crystallization. Indeed, distinct monodispersed species are more prone to crystallize if they are in a steady oligomeric state at a defined concentration.

2. Protein Preparation and Crystallization

1. Preparative SEC
    1. After a one-column volume wash with degassed filtered water, equilibrate with SEC-buffer the preparative pre-packed gel filtration column (a wide range molecular weight 120 mL column).
    2. Inject 1 - 5 mL of purified SBP at 5-50 mg/mL into the equilibrated gel filtration column and flow one column volume of SEC-buffer.
    3. Collect the column flow-through in 1 mL fractions.
    4. Pool the fractions corresponding to a single monodisperse peak; spin the sample in 15 mL centrifugation filters at 1,500 g and quantify the concentration until the desired concentration is attained (typically 50 - 100 mg/mL for SP0092).

2. Crystallization
    1. Determine the optimal concentration range for the crystallization experiment using the Pre-Crystallization Test as described by the manufacturer:
        1. Dispense 0.5 - 1.0 mL of each of the four crystallization test solutions in a different reservoir of a 24-well sitting drop crystallization plate.
        2. Prepare the crystallization drops by mixing 1 µL of protein solution with 1 µL of reservoir solution, seal the plate, and incubate at room temperature for no less than 1 h (more reliable results are obtained after overnight incubation).
        3. Check the drop quality for each concentration using a 10x magnification microscope.
           NOTE: Test at least three different protein concentration: (I) most of the drops remaining clear indicate that the protein is too dilute for crystallization; (II) the presence of heavy precipitate in the majority of the drops implies an excessively high concentration; and (III) a balanced occurrence of both clear drops and precipitate (better if light) is a good indication that the tested concentration is favorable for crystallization.
    2. Prepare the 96-well sitting drop crystallization plates; dispense 100 µL of different commercial crystallization solution in each reservoir well[34,35,36,37].
    3. Dispense 100 nL of protein sample at the previously defined concentration (step 1.1.6) in all the crystallization drop wells using a crystallization robotic system.
    4. Dispense 100 nL of the different reservoir solutions to the corresponding crystallization drop wells to mix with protein dispensed in step 2.2.3. Seal the crystallization plate to avoid evaporation and enable the equilibration of the crystallization drop with the reservoir.
    5. Check the drops periodically (initially every 1 - 2 days, later every week) using a 10x (at least) magnification microscope to evaluate crystal formation and growth.
       NOTE: An automated imaging system for protein crystallization can be used for drops visualization and capturing.

3. Crystal Characterization and X-ray Data Collection

1. Crystal mounting
    1. Prepare the cryoprotectant solution by adding 25% (v/v) of glycerol (final concentration) to the crystallization condition (thus replacing 25% of water in the initial mixture)[38].
    2. Fill a foam dewar with liquid nitrogen and place the sample enclosure part of a uni-puck into the dewar. Allow it to cool to liquid nitrogen temperature.
    3. Cut and remove the sealing tape from the crystallization plate over the drop where the crystals are formed.
    4. Place a drop of 1 µL of cryoprotectant solution onto a coverslide positioned in close proximity of the target drop.
    5. Transfer the selected crystal from the original drop to the cryoprotectant solution drop using a nylon cryo-loop on a SPINE standard base mounted on a magnetic wand[39,40].
    6. Quickly transfer the crystal from the cryoprotectant drop to liquid nitrogen, placing the loop into the first empty position of the uni-puck sample holder.
    7. Repeat (from the sealing tape cut step) until all the desired crystals are harvested and stored in the uni-puck sample holder.
    8. Place the uni-puck base on the uni-puck using the puck wand and take the puck to the beamline (under liquid nitrogen conditions).
       CAUTION: Extremely low temperature!
    9. Use the cryo-tongs and the puck dewar loading tool to load the uni-puck(s) into the beamline sample changer robot dewar. The uni-puck sample holder will detach from the base lid leaving the sample loop holders upright in the sample robot dewar and thus exposed to the liquid nitrogen and accessible to the sample changer robot.

2. Crystal characterization
   NOTE: The harvested crystals can then be screened for diffraction quality using either a laboratory X-ray source or at a synchrotron radiation (SR) X-ray source. Macromolecular crystallography (MX) beamlines provide a tunable energy source to exploit anomalous diffraction for structure solution. In the following section, a series of working instructions in line with the experimental capabilities of Diamond Light Source MX beamlines is proposed, but these guidelines can also be adapted to MX beamlines at other synchrotrons worldwide.
    1. As a first step, it is useful to either confirm or identify anomalous scatterers in the crystal. This can be conveniently achieved by measuring an X-ray fluorescence spectrum from the crystal. First ensure the X-ray energy of the beamline is set high enough to excite all elements typically expected for macromolecular crystals (14 keV or higher).
    2. Use the beamline control software to select the sample to be mounted by the sample changer robot; the loop will be automatically centered in the X-ray beam and the crystal centering can be confirmed manually using the beamline control software.
    3. Record an X-ray fluorescence spectrum using the beamline control software: the user selects an exposure time and initiates the measurement (Fluorescence/Fluorescence Spectrum Setting, → Run, **Figure 2A**). The beamline control software automatically positions the X-ray fluorescence detector in place and determines the minimum incident X-ray flux to obtain a readable signal on the fluorescence detector. The acquired spectrum is then analyzed in an automated fashion with emission peaks fitted to naturally occurring biological elements. These routines are available within the beamline control software graphical user interface (GUI) - GDA (www.opengda.org) at our company and at synchrotrons throughout Europe[41,42].
    4. If suitable elements are identified that can be exploited for an anomalous diffraction experiment, perform an X-ray absorption edge energy scan to determine the optimal wavelength for collection of anomalous diffraction data from the crystal: on the beamline control software, select the element and initiate the scan (Fluorescence/Fluorescence Scan Setting, → Atom name, → Run, **Figure 2B**).
       NOTE: For a single anomalous diffraction experiment, the peak of the absorption edge is used to maximize the anomalous scattering. Experimental considerations for performing a multi-wavelength anomalous diffraction experiment have been previously detailed[43].
    5. To determine the crystal unit cell parameters, symmetry, and diffraction limit, measure three X-ray diffraction patterns at 45° intervals using the oscillation method (Data Collection/Screening → Run current, **Figure 2C**). The beamline control software provides the user with a choice of oscillation angle and exposure time, and the ability to set the percentage transmission of the X-ray beam. For a standard crystal screening, an oscillation angle of 0.5°, an exposure time of 0.5 s, and 5% X-ray beam transmission are recommended.

The measured diffraction images are automatically analyzed by the EDNA pipeline and return a set of strategies for collection of a complete data set[22].

3. X-ray data collection
NOTE: The user can select to collect data in standard oscillation mode or inverse beam mode, which enables Friedel mates to be recorded close in time and X-ray dose, and with approximately the same absorption behavior allowing a more accurate measurement of anomalous differences. The latter is useful especially if small anomalous differences are expected and/or the samples are radiation sensitive when carrying out an anomalous diffraction experiment. Considerations on the best data collection strategies to use have been comprehensively reviewed[22,23,44,45,46,47].
    1. Using the beamline control software, import the data collection parameters as suggested by the strategy program which will provide:
    i. Start position of the ω-rotation axis
    ii. Oscillation width of the ω-axis rotation angle for each diffraction image
    iii. Exposure time for each diffraction image
    iv. Number of images for a complete dataset (indirectly defining the total ω-axis rotation angle for the whole data collection)
    v. Percentage of attenuation of the X-ray beam to avoid over-exposure and radiation damage
    NOTE: At our institution, for running the data collection, the software pipelines for automated processing of the collected data are well established: (i) The diffraction data are reduced (indexed, integrated, and scaled) by the xia2 pipeline, which generates reflection mtz files for the user as input for phasing and structure solution/refinement[48]. (ii) When a significant anomalous signal is detected during data analysis, a first rapid automated structure solution pipeline (fast_ep) using SHELX attempts to solve the heavy atom substructure by experimental phasing, providing a phased electron density map where feasible. A second more comprehensive structure solution pipeline automatically undertakes attempts to solve and build the structure using independent software suites[49,50,51,52]; in cases where this is successful, the user will be provided with an initial model and electron density map. This provides the basis for the user to complete refinement and validate the model with the crystallographic software suites of choice.

Representative Results

This integrated protocol has been proven to be successful with four (two published and two unpublished structures) of six carbohydrate binding protein targets from pneumococcus analyzed to date[32,53]. In this section, we present the biochemical and structural characterization of SP0092 as a representative result to guide structural studies of SBPs in general.

After the SBP SP0092 has been expressed and purified as defined previously[32], the purified protein was analyzed for buffer stability using a thermal shift assay: SP0092 exhibits an increased T_m at pH 6.5 and in the presence of NaCl in the 0 - 0.2 M concentration range (**Figure 3A**). In light of this, the buffer solution for the following steps was defined as: 0.02 M MES pH 6.5, 0.2 M NaCl, 2.5% (v/v) glycerol, 0.5 mM TCEP. The absolute molar mass of the different oligomerization states of SP0092 was measured by MALS coupled to SEC measuring a molecular weight of 187.2, 140.8, 97.0, and 49.4 kDa, corresponding to tetrameric, trimeric, dimeric, and monomeric species, respectively (**Figure 3B**). The analysis of the SEC profile at different protein dilutions revealed that the oligomerization is triggered by increased protein concentration, suggesting that the larger oligomers are more stable at higher concentrations than typically used in crystallization. Indeed, the purified larger oligomeric species directed to crystallization trials, successfully produced protein crystals while the monomer species did not.

The optimized crystals obtained from the native and Se-methionine labeled forms of SP0092 were characterized by X-ray diffraction. Measurement of the X-ray fluorescence from these crystals revealed in both cases emission peaks for Zn being bound to the protein, while Se was detected only for the Se-methionine crystals as expected. Later, X-ray absorption scans at the Se and Zn edges were performed, which provided direct experimental data to tune the incident X-ray wavelength to the respective X-ray absorption edges of either the Zn or Se present in the crystals to maximize the anomalous signal obtainable from the resultant data (**Figure 4A-B**).

After measuring three diffraction patterns at low transmission, a complete anomalous data set was obtained using the data collection strategy suggested by EDNA (**Figure 4C**). The anomalous signal present triggers the automated phasing pipeline at the beamline to determine the sub-structure, and based on the initial experimental phases derived, produces initial maps and models, which can then be refined and validated (**Figure 4D**).

In summary, potential pitfalls of the technique are centered primarily on crystal availability and quality. Optimization of the buffer conditions to improve protein stability as well as identification of the most suitable oligomeric state of the protein to use, when more than one oligomer is identified in solution, can reduce the risk of failure at the crystallization stage. Exploiting the use of bound metal ions identified at an early stage can speed up structure solution and avoid unnecessary production of selenomethionine labeled protein when molecular replacement methods fail.

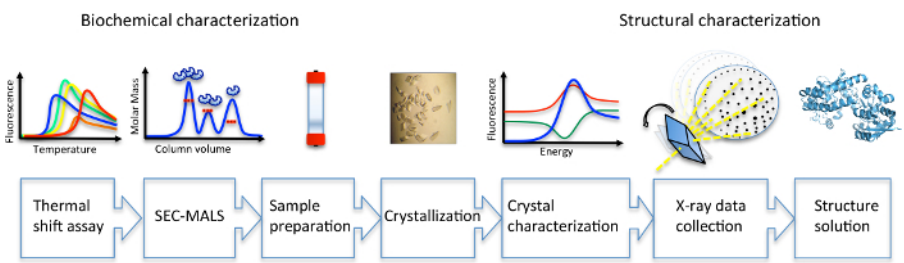
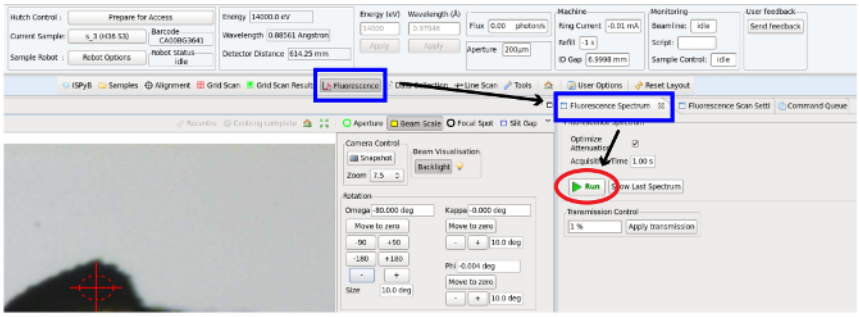
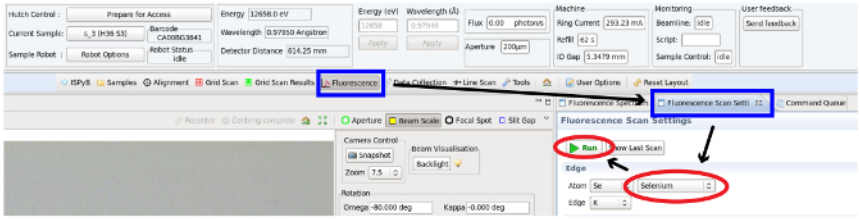


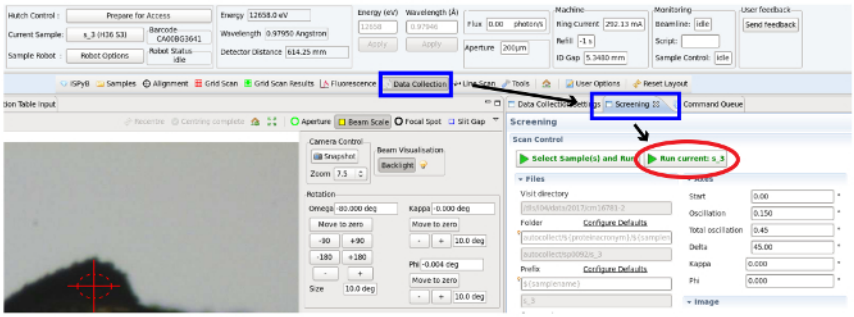
Figure 1. Diagram of workflow for biochemical and structural characterization of carbohydrate SBPs. Please click here to view a larger version of this figure.



A



B



C

Figure 2. Crystal characterization in GDA. (**A**) Screenshot of the X-ray fluorescence control tab of the GDA beamline control software. (**B**) Screenshot of the X-ray energy edge-scan control tab in GDA. (**C**) Screenshot of the data collection crystal screening tab in GDA. Please click here to view a larger version of this figure.

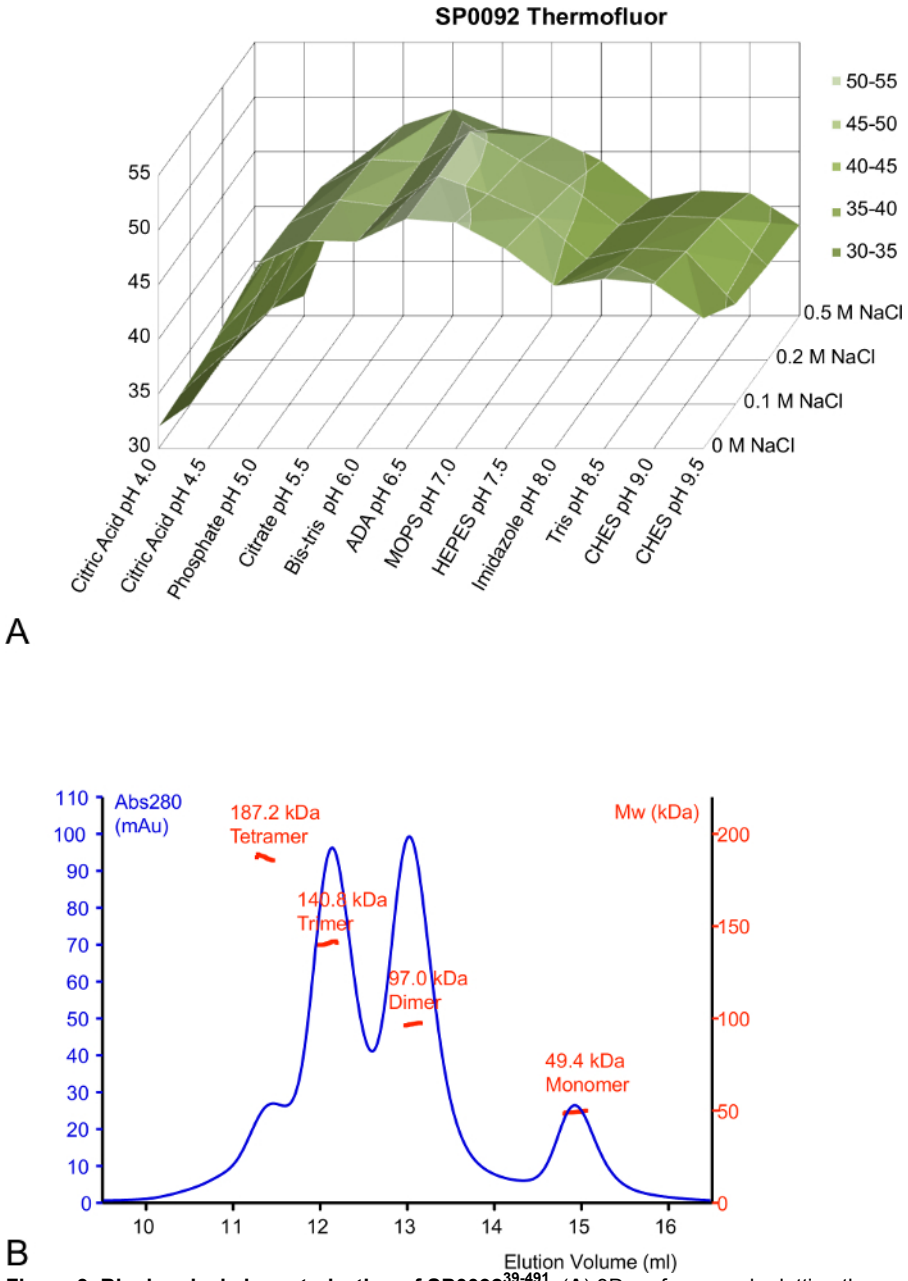


Figure 3. Biochemical characterization of SP0092[39-491]. (**A**) 3D-surface graph plotting the melting temperature of SP0092[39-491] as a function of the pH and NaCl concentration of the buffer solution. (**B**) SEC and MALS results for SP0092[39-491]. Absorption at 280 nm is shown in blue. The molar masses of the different oligomerization states are shown in red. Panel (**B**) has been modified from[32]. Please click here to view a larger version of this figure.

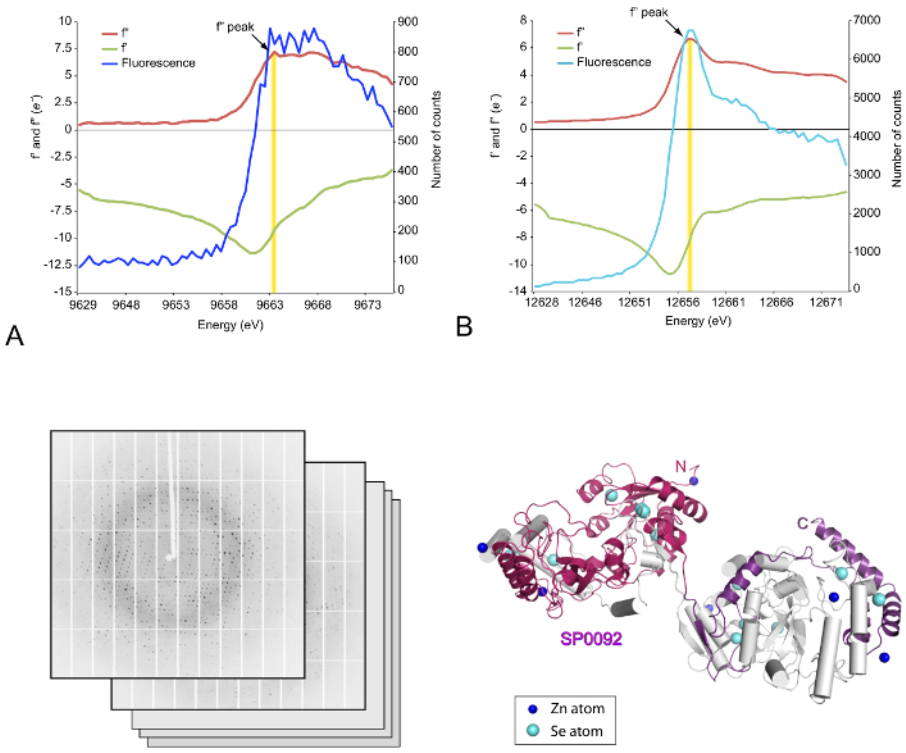


Figure 4. Structural characterization of SP0092$^{39-491}$. (**A**) and (**B**) X-ray absorption energy scan at the Zn and Se edge for SP0092$^{39-491}$ crystals. The measured fluorescence from the Zn and Se containing crystals is shown in blue and cyan, the calculated f' and f'' anomalous scattering fractions are in green and red, respectively. (**C**) Examples of X-ray diffraction patterns collected from SP0092$^{39-491}$ crystals. (**D**) Cartoon representation of the SP0092$^{39-491}$ dimer structure. One protomer is colored white and the other magenta (residues 39 - 366) and violet (residues 367 - 491), and the anomalous scattering atoms are shown as blue and cyan balls for Zn and Se, respectively. Please click here to view a larger version of this figure.

	0.1 M Citric Acid pH 4.0	0.1 M Citric Acid pH 4.5	0.1 M Phosphate pH 5.0	0.1 M Citrate pH 5.5	0.1 M Bis-tris pH 6	0.1 M ADA pH 6.5	0.1 M MOPS pH 7	0.1 M HEPES pH 7.5	0.1 M Imidazole pH 8.0	0.1 M Tris pH 8.5	0.1 M CHES pH 9	0.1 M CHES pH 9.5
NaCl 0 M	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL
NaCl 0.1 M	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL
NaCl 0.2 M	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL
NaCl 0.5 M	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL	40 µL

Table 1. Buffer composition for thermal shift assay.

Discussion

In this paper, we describe and validate an integrated protocol for biochemical and structural characterization of carbohydrate SBPs with a specific emphasis on proteins from *S. pneumoniae*. Nevertheless, this can be used as a standard procedure for the analysis of other SBPs from different organisms and even other unrelated soluble proteins.

The first part of the protocol is focused on providing biochemical information on protein stability and quaternary structure, which can be exploited in the preparation of protein samples for crystallization. In the thermal shift assays section, we describe only the pH and NaCl concentration variations to maintain the general nature of this procedure. Despite this, many other buffer conditions can be tested in a similar way, for example, including any chemical compound used as a stabilizing additive: in particular the actual ligand(s) that bind to a specific SBP are remarkably effective in increasing the T_m by a few degrees Celsius[31]. In some cases, the denaturing curves can be poorly defined due to low signal or a high fluorescence background, which is caused by protein aggregation or partial unfolding. To avoid this, a protein:dye titration can be performed to optimize the unclear denaturing curves. If no improvement is obtained, screening various additives that can ameliorate the stability of the protein is advised, and suitable screens have been previously described[54].

Typically, most SBPs are monomeric in their natural environment, but as shown here multimerization can occur at the higher concentrations used in crystallization experiments, thus the oligomerization behavior characterization provided by MALS and SEC is essential to assess the most favorable stable monodisperse oligomerization state for crystallization. Nevertheless, it is hard to predict the effect of different chemicals included in the crystallization condition on the oligomerization behavior of the proteins. If the SEC and MALS examination shows extensive aggregation of the protein sample, we would advise the following to reduce the likelihood of this occurring: use fresh protein sample (not freeze-thawed) and expand the stabilization analysis performed with thermal shift assays, testing possible additives and mild detergents as a last resource, to minimize aggregation. In this paper, we present basic guidelines for crystallization using high-throughput sparse matrix commercial crystallization screening to maintain the general nature of this protocol. However, obtaining high-resolution X-ray diffraction protein crystals might need iterative fine tuning to optimize the crystallization conditions with respect to precipitant concentration, pH, addition of chemical additives, different temperatures, and other factors changing equilibrium dynamics between the crystal drop and reservoir[16,17].

The second part of the protocol describes the characterization of the protein crystals in order to define the optimal strategy for X-ray diffraction data collection with a specific focus on the acquisition of anomalous data for SAD phasing. Even if SBPs maintain a similar general architecture (and there are many deposited 3D structures potentially usable as starting models), phasing of these proteins by the molecular replacement method is not always straightforward because of the variability of the secondary structure elements and the intrinsic flexibility of these proteins. Hence, we propose the SAD method and highlight that these proteins may already have intrinsically bound metals or indeed non-specific binding of metals from the crystallization buffer conditions, which can provide a range of anomalous diffracting elements as a standard step in our general protocol.

In conclusion, this protocol defines a standard guided workflow of procedures enabling the detailed description of the biochemical and structural features of SBPs that can be exploited to increase the structure determination success rate, as well as accelerate the structural characterization of SBPs in general.

Disclosures

The authors have nothing to disclose.

Acknowledgements

We acknowledge OPPF-UK for assistance in cloning, Gemma Harris for SEC-MALLS and the scientists of beamlines I03 and I04 at Diamond Light Source.

References

1. Austrian, R. Prevention of pneumococcal infection by immunization with capsular polysaccharides of Streptococcus pneumoniae: current status of polyvalent vaccines. *J Infect Dis.* **136 Suppl** S38-42 (1977).
2. Bogaert, D., De Groot, R., Hermans, P. W. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis.* **4** (3), 144-154, (2004).
3. van Mens, S. P. *et al.* Longitudinal analysis of pneumococcal antibodies during community-acquired pneumonia reveals a much higher involvement of Streptococcus pneumoniae than estimated by conventional methods alone. *Clin Vaccine Immunol.* **18** (5), 796-801 (2011).
4. Weycker, D., Strutton, D., Edelsberg, J., Sato, R., Jackson, L. A. Clinical and economic burden of pneumococcal disease in older US adults. *Vaccine.* **28** (31), 4955-4960 (2010).
5. Doern, G. V. Antimicrobial use and the emergence of antimicrobial resistance with Streptococcus pneumoniae in the United States. *Clin Infect Dis.* **33 Suppl 3** S187-192 (2001).
6. Vasoo, S. *et al.* Increasing antibiotic resistance in Streptococcus pneumoniae colonizing children attending day-care centres in Singapore. *Respirology.* **16** (8), 1241-1248 (2011).
7. Black, S. *et al.* Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr Infect Dis J.* **19** (3), 187-195 (2000).
8. Hanage, W. P. *et al.* Diversity and antibiotic resistance among nonvaccine serotypes of Streptococcus pneumoniae carriage isolates in the post-heptavalent conjugate vaccine era. *J Infect Dis.* **195** (3), 347-352, (2007).
9. Burnaugh, A. M., Frantz, L. J., King, S. J. Growth of Streptococcus pneumoniae on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases. *J Bacteriol.* **190** (1), 221-230, (2008).
10. King, S. J. Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? *Mol Oral Microbiol.* **25** (1), 15-24, (2010).
11. Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science.* **293** (5529), 498-506, (2001).
12. Buckwalter, C. M. , King, S. J. Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol.* **20** (11), 517-522, (2012).
13. Bidossi, A. *et al.* A functional genomics approach to establish the complement of carbohydrate transporters in Streptococcus pneumoniae. *PLoS One.* **7** (3), e33320, (2012).
14. Zheng, X., Gan, L., Wang, E., Wang, J. Pocket-based drug design: exploring pocket space. *AAPS J.* **15** (1), 228-241 (2013).
15. Singh, S., Malik, B. K., Sharma, D. K. Molecular drug targets and structure based drug design: A holistic approach. *Bioinformation.* **1** (8), 314-320 (2006).
16. Dessau, M. A., Modis, Y. Protein crystallization for X-ray crystallography. *J Vis Exp.* (47), (2011).
17. Bergfors, T. Protein crystallization: techniques, strategies, and tips. A laboratory manual. International University Line. *Biotechnology Series, La Jolla.* (1999).
18. Scapin, G. Molecular replacement then and now. *Acta Crystallogr D Biol Crystallogr.* **69** (Pt 11), 2266-2275, (2013).
19. Mao, B., Pear, M. R., McCammon, J. A., Quiocho, F. A. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem.* **257** (3), 1131-1133 (1982).

20. McCoy, A. J., Read, R. J. Experimental phasing: best practice and pitfalls. *Acta Crystallogr D Biol Crystallogr.* **66** (Pt 4), 458-469, (2010).
21. Dodson, E. Is it jolly SAD? *Acta Crystallogr D Biol Crystallogr.* **59** (Pt 11), 1958-1965 (2003).
22. Incardona, M. F. *et al.* EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J Synchrotron Radiat.* **16** (Pt 6), 872-879, (2009).
23. Popov, A. N., Bourenkov, G. P. Choice of data-collection parameters based on statistic modelling. *Acta Crystallogr D Biol Crystallogr.* **59** (Pt 7), 1145-1153 (2003).
24. Terwilliger, T. C. *et al.* Can I solve my structure by SAD phasing? Planning an experiment, scaling data and evaluating the useful anomalous correlation and anomalous signal. *Acta Crystallogr D Struct Biol.* **72** (Pt 3), 359-374, (2016).
25. Millan, C., Sammito, M., Uson, I. Macromolecular ab initio phasing enforcing secondary and tertiary structure. *IUCrJ.* **2** (Pt 1), 95-105, (2015).
26. Taylor, G. L. Introduction to phasing. *Acta Crystallogr D Biol Crystallogr.* **66** (Pt 4), 325-338, (2010).
27. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28** (1), 235-242 (2000).
28. Bird, L. E. *et al.* Green fluorescent protein-based expression screening of membrane proteins in Escherichia coli. *J Vis Exp.* (95), e52357 (2015).
29. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35** (6), e45, (2007).
30. Hendrickson, W. A., Horton, J. R., LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9** (5), 1665-1672 (1990).
31. Culurgioni, S., Harris, G., Singh, A. K., King, S. J., Walsh, M. A. Structural Basis for Regulation and Specificity of Fructooligosaccharide Import in Streptococcus pneumoniae. *Structure.* (2016).
32. Culurgioni, S., Tang, M., Walsh, M. A. Structural characterization of the Streptococcus pneumoniae carbohydrate substrate-binding protein SP0092. *Acta Crystallogr F Struct Biol Commun.* **73** (Pt 1), 54-61, (2017).
33. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem.* **357** (2), 289-298, (2006).
34. Wooh, J. W., Kidd, R. D., Martin, J. L., Kobe, B. Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallogr D Biol Crystallogr.* **59** (Pt 4), 769-772 (2003).
35. Gorrec, F. Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. *Drug Discov Today.* **21** (5), 819-825, (2016).
36. Jancarik, J., Scott, W. G., Milligan, D. L., Koshland, D. E., Jr., Kim, S. H. Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of Salmonella typhimurium. *J Mol Biol.* **221** (1), 31-34 (1991).
37. Newman, J. *et al.* Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr.* **61** (Pt 10), 1426-1431, (2005).
38. Garman, E. F., Mitchell, E. P. Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *Journal of Applied Crystallography.* **29** (5), 584-587 (1996).
39. Teng, T.-Y. Mounting of crystals for macromolecular crystallography in a free-standing thin film. *Journal of Applied Crystallography.* **23** (5), 387-391 (1990).
40. Cipriani, F. *et al.* Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* **62** (Pt 10), 1251-1259, (2006).
41. Gabadinho, J. *et al.* MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *J Synchrotron Radiat.* **17** (5), 700-707, (2010).
42. Leonard, G. A. *et al.* Online collection and analysis of X-ray fluorescence spectra on the macromolecular crystallography beamlines of the ESRF. *Journal of Applied Crystallography.* **42** (2), 333-335 (2009).
43. Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I., Joachimiak, A. MAD data collection - current trends. *Acta Crystallogr D Biol Crystallogr.* **55** (Pt 10), 1726-1732 (1999).
44. Dauter, Z. Data-collection strategies. *Acta Crystallogr D Biol Crystallogr.* **55** (Pt 10), 1703-1717 (1999).
45. Finke, A. D. *et al.* Advanced Crystallographic Data Collection Protocols for Experimental Phasing. *Methods Mol Biol.* **1320** 175-191 (2016).
46. Gonzalez, A. Optimizing data collection for structure determination. *Acta Crystallogr D Biol Crystallogr.* **59** (Pt 11), 1935-1942 (2003).
47. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr.* **62** (Pt 1), 48-57, (2006).
48. Winter, G., Lobley, C. M., Prince, S. M. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr.* **69** (Pt 7), 1260-1273, (2013).
49. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr.* **66** (Pt 4), 479-485, (2010).
50. Skubak, P., Pannu, N. S. Automatic protein structure solution from weak X-ray data. *Nat Commun.* **4** 2777 (2013).
51. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr.* **65** (Pt 6), 582-601, (2009).
52. Vonrhein, C., Blanc, E., Roversi, P., Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol Biol.* **364** 215-230, (2007).
53. Culurgioni, S., Harris, G., Singh, A. K., King, S. J., Walsh, M. A. Structural Basis for Regulation and Specificity of Fructooligosaccharide Import in Streptococcus pneumoniae. *Structure.* **25** (1), 79-93, (2017).
54. Boivin, S., Kozak, S., Meijers, R. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification.* **91** (2), 192-206 (2013).